CHAPTER 3 COST DATABASE

3.1 GENERAL

The cost database is the backbone of the effort to obtain typical costs for the seismic rehabilitation of buildings. This chapter discusses the methods used in collecting and sorting the data including acceptance/rejection procedures and other quality control processes. The data points in the database for this report are either actual construction costs or costs from detailed seismic rehabilitation studies.

3.2 DATA COLLECTION PROCESS

The process of collecting data for this study was developed so as to be as objective as possible. The strength of the database is intended to be its consistency regardless of the person or firm submitting data, the location and date of study of the projects examined, and the types of buildings and performance objectives selected.

The Data Collection Guidelines, as the two-page worksheet that guided the data collection effort is called, requests a broad range of information on a given project. Appendix A contains a copy of this worksheet and the list of data collected. The building framing, layout and codes used in the rehabilitation were obtained to assist in the quality control check. When critical information (area, costs, building type, NEHRP map seismic area, year of study, and performance objective) was unavailable, the worksheets were not added to the database. Where other information was missing the record was assumed to have a lower level of accuracy than those which were complete.

The cost basis was developed as follows:

- **Step One: Identification of Sources of Data**
  Lists of engineers and others familiar with seismic rehabilitation work were gathered. All members of the Advisory Panel were required to provide information on

rehabilitation projects. Firms and individuals on the lists were contacted, the project explained in brief and their help requested in collecting the data.

- **Step Two: Collect Data from First Edition Database**
The second step of the cost data collection was to examine the data which had been collected for the First Edition of the Typical Costs FEMA study done in 1988. While this data was generally much less complete than the newer information, approximately 60% of it was used in the new database because it was examined and found to be acceptable, especially for URM buildings.

- **Step Three: Collect New Data**
The individuals identified in Step One were contacted and the worksheets on the various projects were completed.

- **Step Four: Quality/Data**
Once the completed worksheets were collected, a careful process of quality assurance was undertaken. If necessary information was missing, the person who filled out the worksheet was contacted to help fill in any blanks. Costs were also checked to verify that non-structural costs were properly separated from structural costs.

- **Step Five: Enter Costs into Database**
The information was entered into the database, after each worksheet was thoroughly reviewed for completeness and accuracy.

3.3 TIME AND LOCATION COST ADJUSTMENTS

Much of the information collected was from studies or construction done before 1993. To be consistent, all cost data in the database was indexed to March 1993. For this adjustment of cost the Engineering News Record's (ENR) 20-city average of building costs, called the Building Cost Index (BCI), which compares the historical costs of selected materials and labor to today's costs was used.

For costs associated with studies done before 1970, the index factor rises rapidly and for this time period the cost correction was done in consultation with Hanscomb Associates, a member of the Advisory Panel.

In addition to indexing the data based on the year of the study or construction year, costs from various parts of the country and Canada were referenced to the St. Louis location, to account for regional differences in labor and material rates. To account for these differences another correction was made to each cost data point. The Means Index relates costs in 250 cities in the United States and Canada. For each state, U.S. territory or Canadian province where data was collected, an average factor of all the cities in the state, territory or province was calculated and compared to the common location, which was chosen as Missouri. Missouri was selected to be the baseline state for this study solely because of its central geographic location. Thus, where all cities in Missouri were given a baseline of 1.00, all buildings in South Carolina, for example, were factored by 0.80. Canadian factors took into account the 1993 average exchange rate so that Canadian dollar amounts entered on the work sheets for buildings in Canada could be directly converted to U.S. dollars.

The factors correcting for the year of construction or study and the location factors were multiplied together to obtain a combined factor. All costs for each building were multiplied by the appropriate factor so that each building cost is relative to March, 1993 in Missouri dollars.

3.4 DATA QUALITY RATING

There is a notable variation in the quality of the cost data. The project goal was to not eliminate any data except that which lacked enough minimum information to be useful. Therefore, each cost data point was assigned a quality rating. Quality factors were calculated for each building cost data value, ranging from 1 (being the least accurate) to 10 (being the most accurate).

Care was taken to make the rating system as objective as possible so that another uncertainty, that of the engineer assigning the factor, would be minimized. The rating was determined as the sum of the following three parameters:

- **Date of study:** Design professionals today are more familiar with earthquakes, seismic rehabilitation methods and building performance. Consequently, the accuracy of their cost estimates has increased considerably. Therefore, the rating in Table 3.4.1 was given to each record based on the date of its cost study or construction.

TABLE 3.4.1 QUALITY/RATING DATE OF STUDY

DATE OF STUDY OR CONSTRUCTION	POINTS
Before 1973	1
Between 1973 and 1987	2
After 1987	3

- **Source and certainty of cost:** Each design professional was asked to check whether the cost estimate on the Data Collection Guidelines was from a study or actual construction. Also, the design professional rated his or her confidence in the costs as either Good, Fair or Poor. Based on these choices, the ratings in Table 3.4.2 were given.

TABLE 3.4.2 QUALITY RATING/SOURCE AND CERTAINTY OF COST

SOURCE	CONFIDENCE	POINTS
Unknown	Poor	0
Study	Poor	1
Study	Fair or Good	2
Actual	Poor	2
Actual	Fair	3
Actual	Good	4

- **Consistency of data:** In many instances the information provided for particular buildings or groups of buildings was sporadic and incomplete. Older or general studies of large numbers of buildings often contained minimal information. The familiarity and experience with seismic rehabilitation of the person filling out the worksheet would, in general, affect the quality of the data. So that no single characteristic would weigh too heavily on the point value given to this factor, the following procedure was used: seven characteristics were developed by which each record would be rated, with a 1 (positive) or a 0 (unknown or negative). These characteristics were: Were the worksheets complete and clearly filled out? Did the person or office submit many records or only a few? Were the reports from which the worksheets were prepared specific and complete? Was the engineer located in a region of high seismicity? Was the person or office submitting the forms a member of the Advisory Panel? Was the person filling out the worksheets a registered Structural Engineer or Architect? Was the person or firm submitting the information well recognized in the earthquake engineering profession?

Based on the total point value obtained from this list of characteristics, a rating was given for the consistency parameter as shown in Table 3.4.3:

TABLE 3.4.3 QUALITY RATING/CONSISTENCY OF DATA

SUM OF CHARACTERISTICS	POINTS
0-1	0
2-3	1
4-5	2
6-7	3

Figure 3.4.1 shows the number of buildings versus the quality rating for the three categories of the performance objective. Figure 3.4.2 shows the same plot as a function of the seismicity.

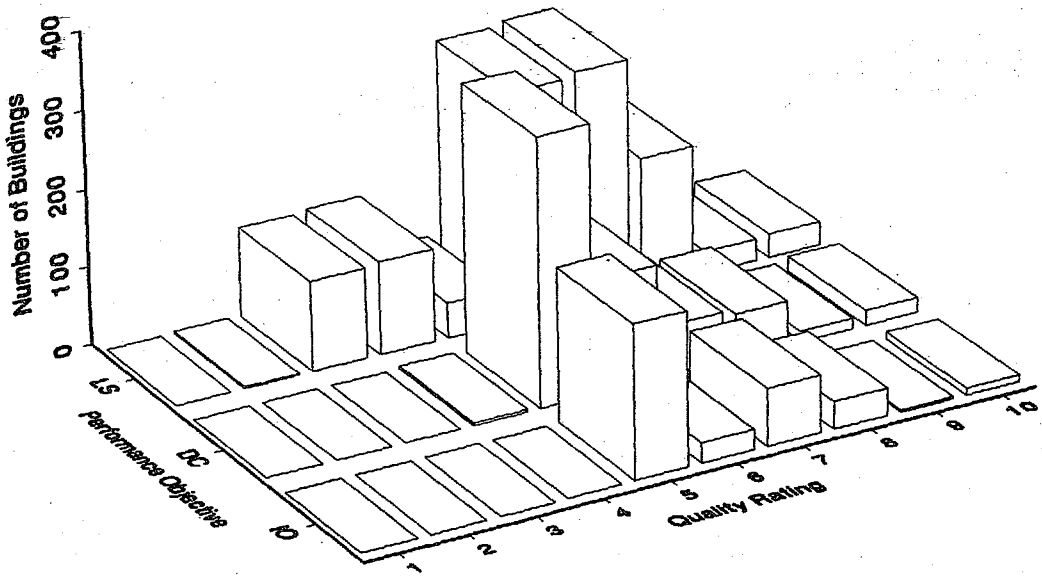

FIGURE 3.4.1 NUMBER OF BUILDINGS FOR DIFFERENT QUALITY RATING/
PERFORMANCE OBJECTIVE COMBINATIONS
(WITHOUT URM BUILDINGS)

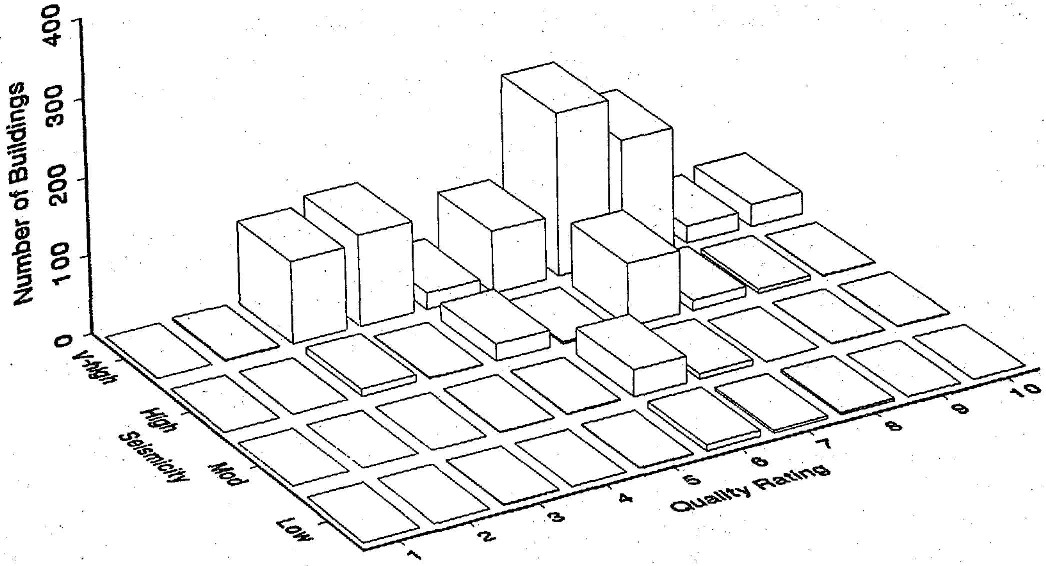

FIGURE 3.4.2 NUMBER OF BUILDINGS FOR DIFFERENT QUALITY RATING/
SEISMICITY COMBINATIONS
(LIFE SAFETY PERFORMANCE OBJECTIVE)

3.5 SUPER DATABASE

The database that was obtained by using the process described earlier contained 2088 cost data points and could have been directly used to develop the cost estimation coefficients in the methodology that is presented in Chapter 4. However, if that procedure had been followed, it would have not taken advantage of the information about the difference in quality between the cost data points as described and quantified in Section 3.4. Therefore, a super cost database was developed using the 2088 cost data values and their associated quality rating and a weighting process than incorporates the relative value of the cost data and the confidence in the value of that cost data.

The super database was developed by taking each of the original 2088 cost data points and, one at a time, using them to generate several new values of cost. For each original cost data value, the number of new cost values that go into the super database is a function of the quality rating of that data value, see Figure 3.5.1. For example, if the quality rating was 7, then 83 new cost data points would go into the super database.

Similarly, if the quality rating was 5 and not 7, then only 72 new cost data points would go into the super database. Therefore, the super database will contain more data for the higher quality rating. The value of each of the new cost data points that goes into the super database incorporates the increased confidence in the value of the cost that is associated with the higher quality rating of the data. Each new cost data point that was created for the super database was generated using a Monte Carlo Simulation Analysis (MCS) using an underlying lognormal probability distribution with a mean sample value equal to the cost of the original data point and a coefficient of variation related to the quality rating, see Figure 3.5.2. Repeating this for all original data points results in the super cost database that is used to perform the analysis that yields the cost estimation equations in Chapter 4. The details of this database generation are given in Volume 2.

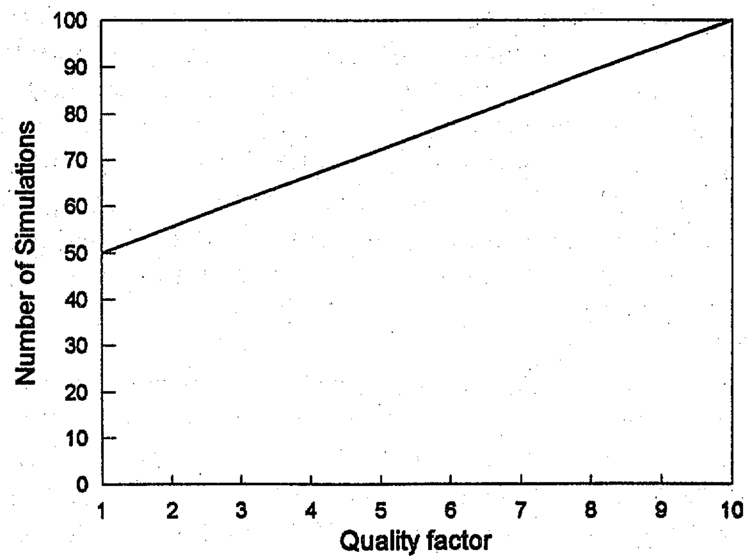

FIGURE 3.5.1 NUMBER OF SIMULATIONS FOR NEW COST DATA

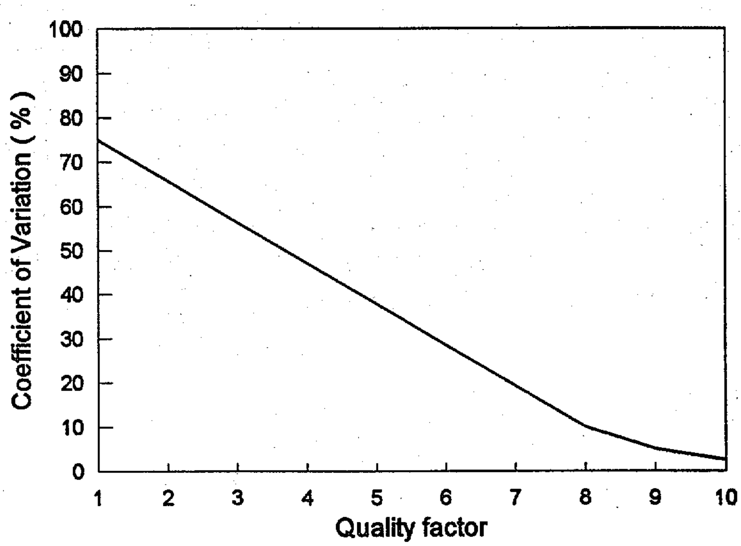

FIGURE 3.5.2 COEFFICIENT OF VARIATION FOR NEW COST DATA